

Austin L. Hughes · Bernice Packer · Robert Welch ·
Stephen J. Chanock · Meredith Yeager

High level of functional polymorphism indicates a unique role of natural selection at human immune system loci

Received: 21 July 2005 / Accepted: 23 September 2005 / Published online: 29 October 2005
© Springer-Verlag 2005

Abstract Several studies have shown that immune system proteins have on average a higher rate of amino acid evolution between different species of mammals than do most other proteins. To test whether immune-system-expressed loci show a correspondingly elevated rate of within-species nonsynonymous (amino acid altering) polymorphism, we examined gene diversity (heterozygosity) at 4,911 single nucleotide polymorphism (SNP) sites at 481 protein-coding loci. At loci with nonimmune functions, gene diversity at nonsynonymous SNP sites was typically lower than that at silent SNP sites (those not altering the amino acid sequence) in the same gene, a pattern that is an evidence of purifying selection acting to eliminate slightly deleterious variants. However, this pattern was not seen at nonsynonymous SNPs causing conservative amino acid replacements in immune system proteins, indicating that the latter are subject to a reduced level of functional constraint. Similarly, immune system genes showed higher gene diversities in their 5'

noncoding regions than did other proteins. These results identified certain immune system loci that are likely to be subject to balancing selection that acts to maintain polymorphism in either coding or regulatory regions.

Keywords Gene diversity · Interleukin · Natural selection · Single nucleotide polymorphism

Introduction

The genes of the vertebrate immune system have provided several well-studied cases of positive Darwinian selection at the molecular level. For example, there are several lines of evidence supporting the hypothesis that the loci of the class I and class II major histocompatibility complex (MHC) are subject to overdominant selection or some similar form of balancing selection (Hughes and Yeager 1998). The MHC genes are characterized by an unusual pattern of nucleotide substitution in the peptide-binding region (PBR) of the molecule, where the number of nonsynonymous nucleotide substitutions per nonsynonymous site (d_N) exceeds the number of synonymous substitutions per synonymous site (d_S) (Hughes and Nei 1988, 1989). This pattern of nucleotide substitution is evidence that natural selection has acted to favor amino acid sequence polymorphism in the PBR (Hughes and Nei 1988). Furthermore, there are several well-documented cases of directional selection leading to amino acid sequence diversification within families of immune system genes, as indicated by d_N exceeding d_S in comparisons between paralogs. Examples include immunoglobulin V region genes (Tanaka and Nei 1989), killer cell immunoglobulin-like receptor (Hughes 2002), and defensins (Hughes 1999a).

In addition to the evidence of positive selection acting on individual immune system gene families, there is evidence that immune system genes of mammals are on average subject to less stringent purifying selection than most other genes. Murphy (1993) compared 615 proteins encoded by putatively orthologous genes of human and murine rodents and found a higher mean rate of amino acid sequence

Electronic Supplementary Material Supplementary material is available for this article at <http://dx.doi.org/10.1007/s00251-005-0052-7>.

A. L. Hughes (✉)
Department of Biological Sciences,
University of South Carolina,
Coker Life Sciences Bldg., 700 Sumter St.,
Columbia, SC 29208, USA
e-mail: austin@biol.sc.edu

B. Packer · R. Welch · M. Yeager
Intramural Research Support Program,
SAIC-Frederick, NCI-FCRDC,
Frederick, MD 21702, USA

B. Packer · R. Welch · S. J. Chanock · M. Yeager
Core Genotyping Facility,
Division of Cancer Epidemiology and Genetics,
National Cancer Institute, National Institutes of Health,
Bethesda, MD 20892-4605, USA

S. J. Chanock
Section on Genomic Variation, Pediatric Oncology Branch,
National Cancer Institute, National Institutes of Health,
Gaithersburg, MD 20892-4605, USA

evolution in host defense ligands and receptors than in any other functional category of proteins. Likewise, a comparison of immunoglobulin superfamily C2 type domains from mammalian proteins expressed in immune system cells and other tissues showed evidence of a higher rate of nonsynonymous substitution in the former than in the latter (Hughes 1997). The increased rate of amino acid evolution at immune system loci might reflect either a relaxation of functional constraint on immune system proteins in comparison to other proteins or the cumulative effect of numerous events of positive selection, perhaps driven by a coevolutionary race with parasites (Murphy 1993; Hughes 1997, 1999b).

In spite of the evidence of an enhanced rate of amino acid evolution at immune system loci over the history of placental mammals, so far, it is unknown whether there is a corresponding enhancement of nonsynonymous polymorphism at these loci within species. Here, we address this question using data on 4,911 single nucleotide polymorphisms (SNPs) at 481 protein-coding loci from the SNP500 database, a database that provides high-reliability typing of SNPs on a reference population of 102 individuals representing a sampling of worldwide human genetic diversity (Packer et al. 2004). The SNPs included in this database were chosen because they are found within known protein-coding genes involved in fundamental biological processes (including development, the cell cycle, and immunity) believed to be relevant to cancer and other complex human diseases (Packer et al. 2004). Because the focus of the database is on SNPs of potential functional relevance, the SNPs included were heavily biased toward exons, intron-exon borders, and regions within 5 kb of the start or end of the open reading frames.

Previous analysis of smaller data sets from the same database showed that gene diversity (heterozygosity) was significantly lower at SNP sites where one allele caused a radical amino acid change or introduced a stop codon than at silent SNP sites in the same genes (Hughes et al. 2003, 2005). These results, like similar results from other studies (Freudenberg-Hua et al. 2003; Sunyaev et al. 2003; Zhao et al. 2003), support the hypothesis that the human population includes a large number of slightly deleterious mutations which are subject to ongoing purifying selection (Hughes et al. 2003). Here, we compare SNP at immune system genes with that at other genes to determine whether natural selection acts similarly at the former and the latter.

Methods

Samples

Allele frequency data for 4,911 SNP sites at 481 loci (Supplementary Table S1) were taken from the SNP500 database (Packer et al. 2004). This database is based on bidirectional sequence determination of each SNP in 102 unrelated individuals from the Coriell Institute for Medical Research (Camden, NJ, USA; <http://locus.umdnj.edu/nigms/>). On the basis of self-described ethnicity, these individuals were assigned to four groups: 31 non-Hispanic

Caucasians, 24 Africans and African-Americans, 24 of Pacific Rim heritage, and 23 Hispanics. This population thus cannot be considered a random sample of the worldwide human population, but it samples worldwide human genetic variation by including representatives of the major geographic subdivisions of the worldwide human population. Three of the four groups (excluding Hispanics) correspond to the three major historic subdivisions of the human population: Africans, Europeans, and Asians (Hughes et al. 2005). For details of sequencing methodology, see Packer et al. (2004) and Hughes et al. (2003). Allelic frequency data for all SNPs are available at <http://snp500cancer.nci.nih.gov>.

Statistical analyses

Gene diversity (heterozygosity) at a SNP site was estimated by $1 - \sum_{i=1}^n x_i^2$ where n is the number of alleles and x_i is the population frequency of the i th allele (Nei 1987, p. 177). SNPs were classified with respect to their location and effect on protein function as follows: (1) SNPs in the 5' noncoding region, which included both the 5' UTR and the 5' intergenic region (within 10,000 bp of the start codon); (2) SNPs in the 3' noncoding region; (3) SNPs in introns; (4) synonymous SNPs in exons; (5) nonsynonymous SNPs in exons causing a conservative amino acid change; and (6) nonsynonymous SNPs in exons causing a radical change. Radical nonsynonymous changes included a small number of cases ($N=5$) where an SNP introduced a stop codon, and a larger number of cases causing an amino acid replacement involved two amino acids with a pairwise stereochemical difference >3.0 according to the scale of Miyata et al. (1979) (based on amino acid residue volume and polarity). Otherwise, missense SNPs were categorized as conservative. The scale of Miyata et al. (1979) is one of the many scales that measure chemical distance between amino acids, all of which are positively correlated with one another. Use of other scales in preliminary analyses yielded similar results to those using the scale of Miyata et al. (1979). Previous analyses of gene diversity at a smaller number of SNP sites from the same population showed a similar pattern in the case of nonsense SNPs and radical nonsynonymous SNPs (Hughes 2003); therefore, the two were combined in a single category in the present analyses. All types of SNPs except nonsynonymous SNPs were classified together as "silent" SNPs.

On the basis of descriptions of protein function in the biological literature, loci were placed in three sets: (1) the immune set (65 loci), including loci encoding proteins with a role in processes of specific and innate immunity; (2) the hormone set, including loci encoding hormones and growth factors and their receptors (50 loci); and (3) the remainder set, including all other loci (366 loci) (Supplementary Table S1). The immune set did not include any of the highly polymorphic loci of the class I and class II MHC. Functional classification was straightforward because the SNP500 loci were chosen originally because of

their well-described functions (Packer et al. 2004). Hormones and growth factors and their receptors were analyzed separately to test whether patterns observed in the immune system set reflected immune function or merely characteristics of signaling molecules and their receptors. The hormone, consisting of nonimmune signaling molecules and receptors, thus represented a control group in testing the hypothesis that natural selection acts in a unique way on immune system genes. In addition, Murphy (1993) reported a relatively high rate of amino acid evolution between human and rodent in hormones and their receptors.

For pairwise estimates of linkage disequilibrium (LD) among SNPs, haplotypes were estimated from unphased genotypes using the PHASE program (Stephens et al. 2001). We used Haplovview (Barrett et al. 2005) to estimate D' (the LD expressed as a proportion of the maximum disequilibrium possible for a given set of allele frequencies) and r^2 (the correlation between two sites).

The distribution of gene diversities at individual SNP sites was positively skewed and differed significantly from normality (Kolmogorov–Smirnov test $P<0.001$). Therefore, nonparametric statistical tests were used to compare median gene diversities at individual SNP sites among different SNP functional categories and different sets of loci. To compare gene diversity and radical nonsynonymous SNP sites and conservative nonsynonymous SNP sites with that of silent sites in the same genes, we computed within-gene means of gene diversity at each class of nonsynonymous SNP and compared the means pairwise with mean gene diversity at silent sites in the same gene. We used pairwise t tests for this comparison, since the distribution of pairwise differences was not significantly different from normality (Kolmogorov–Smirnov test).

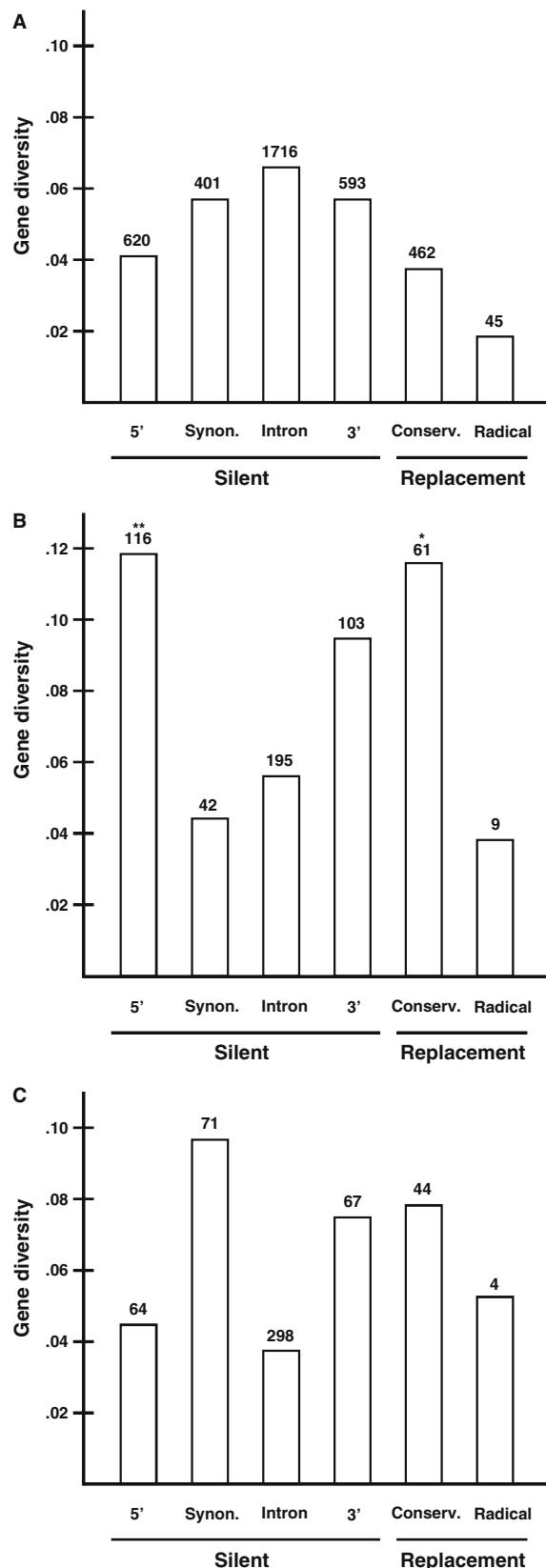
Results

Differences among single nucleotide polymorphism functional categories

SNP sites (4,911) at 481 protein-coding loci were classified in three sets based on the function of the encoded: an immune set (immune system proteins), a hormone set (hormones and growth factors and their receptors), and a remainder set (all others). SNPs in the remainder set showed a pattern of gene diversity similar to that reported previously (Hughes et al. 2003, 2005) (Fig. 1a). Median gene diversity differed significantly ($P<0.001$, Kruskal–Wallis test) among the six functional categories of SNPs, with substantially lower median gene diversity (0.019) at radical nonsynonymous SNP sites than at sites in other categories (Fig. 1a). In addition, median gene diversity at the conservative nonsynonymous SNP sites (0.038) and at sites in 5'

Fig. 1 Median gene diversity at SNP sites for functionally distinct SNP categories: **a** remainder set (medians significantly different among categories; $P<0.001$, Kruskal–Wallis test), **b** immune set, and **c** hormone set. Median significantly different from that for corresponding category of remainder set: * $P<0.05$ ** $P<0.01$

noncoding regions (0.041) was slightly lower than those at sites in 3' noncoding regions (0.057), at sites in introns (0.067), or at synonymous sites in exons (0.059).



By contrast, neither the immune set (Fig. 1b) nor the hormone set (Fig. 1c) showed a significant difference in median gene diversity among SNP functional categories. However, SNPs in the 5' noncoding region showed a significantly higher median gene diversity in the immune set than in the remainder set (Wilcoxon rank sum test, $P<0.01$; Fig. 1b). Similarly, conservative nonsynonymous SNPs showed a significantly higher median gene diversity in the immune set than in the remainder set (Wilcoxon rank sum test, $P<0.05$; Fig. 1c).

Conservative nonsynonymous SNPs in the immune set included proportionately fewer sites with low gene diversity and proportionately more sites with high gene diversity than did the remainder set (Fig. 2). Of 61 conservative nonsynonymous SNPs in the immune set, 37 (60.7%) had gene diversities greater than 0.05, whereas only 211 of 462 (45.3%) conservative nonsynonymous SNPs in the remainder data set had gene diversities greater than 0.05. The difference in proportions between the two sets was statistically significant (chi-square test, $P=0.028$). Similarly, 70 of 116 (60.3%) 5' noncoding SNPs in the immune set had gene diversities greater than 0.05, whereas only 293 of 620 (47.3%) of 5' noncoding SNPs in the remainder set had gene diversities greater than 0.05. Again, the difference in proportions between the two sets was statistically significant (chi-square test, $P=0.01$).

Table 1 lists conservative nonsynonymous and 5' non-coding SNP sites in the immune set with exceptionally high gene diversity (>0.40). It is of interest that certain genes included more than one such high-diversity SNP site. For example, there were three high-diversity SNP sites in the 5' noncoding region of the *IL1B* gene and, likewise, three high-diversity SNP sites in the 5' noncoding region of the *IL10* gene (Table 1). At the *CSF2*, *CTL4A*, and *IL4R* loci, there was a high-diversity SNP site in the 5' noncoding region and another at a conservative nonsynonymous site in the coding region.

Within-locus comparisons

The reduction of gene diversity at nonsynonymous SNP sites, in comparison to other SNP sites in the same genes, is the evidence of purifying selection on those sites (Hughes et al. 2003, 2005). Conversely, the absence of such a reduction in the immune set suggests a relaxation of purifying selection. To investigate further natural selection on nonsynonymous SNPs in the immune, hormone, and remainder sets, we compared within each protein-coding locus the mean gene diversity at nonsynonymous SNPs with mean gene diversity at silent SNP sites (all SNP sites in the 5' and 3' noncoding regions, SNP sites in introns, and synonymous SNP sites in exons). We compared the means of the per-locus gene diversities at both radical and conservative nonsynonymous SNP sites pairwise with those at silent SNP sites in the same genes (Table 2). In the remainder set, mean gene diversities for both radical and

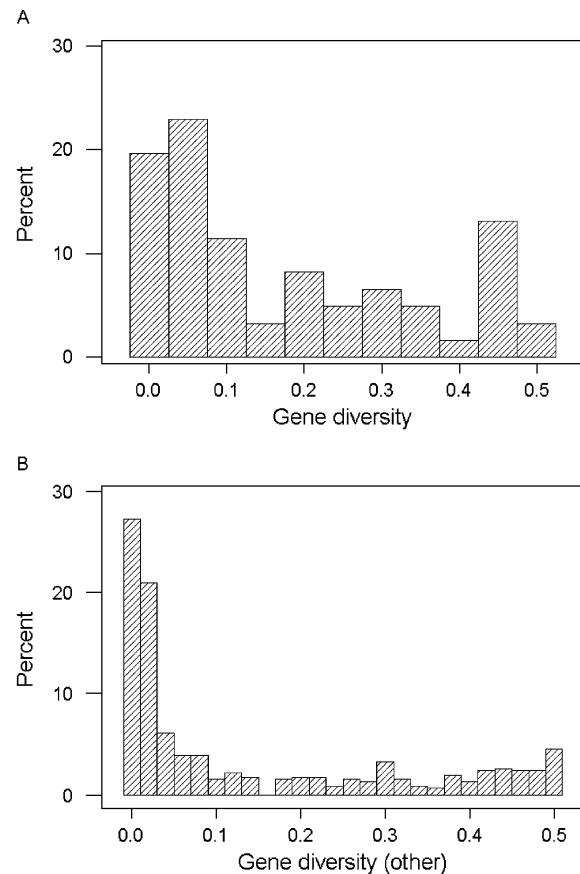


Fig. 2 Distribution of gene diversity values at conservative nonsynonymous sites in **a** immune set and **b** remainder set. The two sets differed significantly with respect to the proportion of SNPs with gene diversity greater than 0.05 (chi-square test, $P=0.012$)

conservative SNPs were significantly lower than those for silent sites in the same genes (Table 2).

By contrast, in both the immune set, the mean gene diversity of radical SNPs was significantly lower than that for silent SNPs, but the mean gene diversity of conservative nonsynonymous SNPs was not significantly different from that for silent SNPs (Table 2). In the hormone set, neither radical nor conservative nonsynonymous SNPs showed gene diversities significantly different from those of silent SNPs in the same genes (Table 2). Thus, while both radical and conservative SNPs in the remainder set showed evidence of purifying selection reducing their gene diversities below those of silent SNPs in the same genes, this evidence was lacking for conservative nonsynonymous SNPs in the immune and hormone sets.

When we plotted mean gene diversity at conservative nonsynonymous SNPs against mean gene diversities at silent SNPs in the immune set, we found that these quantities were in general not related (Fig. 3). Certain loci showed high-mean gene diversity at conservative nonsynonymous SNPs and low-mean gene diversities at silent nonsynonymous SNPs, while others showed the

Table 1 SNP sites in 5' noncoding regions or conservative nonsynonymous sites in exons of immune system loci with high (>0.40) gene diversity

SNP category	Locus	Protein encoded	SNP ID ^a	Gene diversity
5' noncoding	<i>CCL5</i>	Chemokine (C-C motif) ligand 5	rs2107538	0.4596
	<i>CSF2</i>	Colony-stimulating factor 2	rs1469149	0.4051
	<i>CTLA4</i>	Cytotoxic T-lymphocyte-associated protein 4	rs11393378	0.4838
	<i>IFNG</i>	Interferon gamma	rs2069705	0.5000
	<i>IL1A</i>	Interleukin-1 alpha	rs1800587	0.4100
	<i>IL1B</i>	Interleukin-1 beta	rs1143623	0.4571
	<i>IL1B</i>	Interleukin-1 beta	rs1143627	0.4999
	<i>IL1B</i>	Interleukin-1 beta	rs16944	0.4992
	<i>IL4</i>	Interleukin-4	rs2070874	0.4510
	<i>IL4</i>	Interleukin-4	rs2243250	0.4835
	<i>IL4R</i>	Interleukin-4 receptor	rs2107356	0.4051
	<i>IL5</i>	Interleukin-5	rs2069812	0.4917
	<i>IL8</i>	Interleukin-8	rs4073	0.4969
	<i>IL10</i>	Interleukin-10	rs1800871	0.4788
	<i>IL10</i>	Interleukin-10	rs1800872	0.4816
	<i>IL10</i>	Interleukin-10	rs1800896	0.4152
	<i>IL15</i>	Interleukin-15	rs1493012	0.4192
	<i>IL15</i>	Interleukin-15	rs1493013	0.4192
	<i>LTA</i>	Lymphotoxin alpha	rs2239704	0.4562
	<i>MBL2</i>	Mannose-binding lectin 2	rs11003125	0.4444
	<i>TLR2</i>	Toll-like receptor 2	rs4696480	0.4861
	<i>TNFRSF6</i>	Tumor necrosis factor receptor superfamily, member 6	rs1800682	0.4904
Conservative nonsynonymous	<i>CD86</i>	CD86 antigen (CD28 antigen ligand 2, B7-2 antigen)	rs1129055	0.4688
	<i>CFH</i>	Complement factor H	rs800292	0.4675
	<i>CSF2</i>	Colony-stimulating factor 2	rs25882	0.4477
	<i>CTLA4</i>	Cytotoxic T-lymphocyte-associated protein 4	rs231775	0.4142
	<i>FCGR2A</i>	Fc fragment of IgG low-affinity IIa receptor	rs1801274	0.4919
	<i>ICAM1</i>	Intercellular adhesion molecule 1	rs5030382	0.4498
	<i>IFNAR2</i>	Interferon receptor 2	rs1051393	0.4616
	<i>IL4R</i>	Interleukin-4 receptor	rs1801275	0.4694
	<i>IL6R</i>	Interleukin-6 receptor	rs8192284	0.4302
	<i>IL7R</i>	Interleukin-7 receptor	rs1494555	0.4365
	<i>TNFRSF10A</i>	Tumor necrosis factor receptor superfamily, member 10a	rs4871857	0.4855

^aID number in the public SNP database, dbSNP (Sherry et al. 2001)

opposite pattern (Fig. 3). However, two loci stood out in having high-mean gene diversity in both conservative nonsynonymous and silent SNPs: *CFH* and *CSF2* (Fig. 3). In the case of *CFH*, the data set included a single conservative nonsynonymous SNP site (Table 1) and

a single SNP site in an intron, both with high gene diversity. Similarly, the data for *CSF2* included a single conservative nonsynonymous SNP site (Table 1) and a single SNP site in the 5' noncoding region, both with high gene diversity.

Table 2 Within-locus mean gene diversities (\pm SE) for radical and conservative nonsynonymous SNPs compared with those for silent SNPs in the same genes

Locus set	No. of loci	Radical nonsynonymous	Silent	P (paired <i>t</i> test)
Immune	5	0.050 \pm 0.015	0.177 \pm 0.047	0.048
Hormone	4	0.129 \pm 0.093	0.157 \pm 0.085	NS
Remainder	37	0.074 \pm 0.018	0.178 \pm 0.013	<0.001
Locus set	No. of loci	Conservative nonsynonymous	Silent	P (paired <i>t</i> test)
Immune	28	0.184 \pm 0.026	0.172 \pm 0.021	NS
Hormone	18	0.205 \pm 0.043	0.150 \pm 0.024	NS
Remainder	168	0.150 \pm 0.010	0.180 \pm 0.007	0.014

NS Not significant

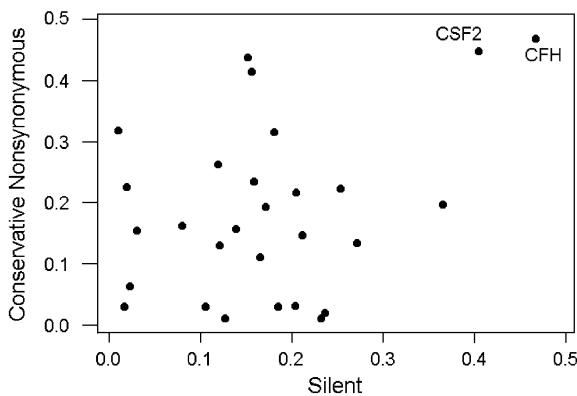


Fig. 3 Plot of mean gene diversity at conservative nonsynonymous sites vs mean gene diversity in all silent sites for 29 immune set loci ($r=0.467$, $P=0.011$). Loci with unusually high values of both variables are indicated

Linkage disequilibrium

Using haplotypes estimated by the PHASE program (Stephens et al. 2001), we analyzed LD among SNP sites at the *IL1B* and *IL10* loci, each of which had three SNPs in the 5' noncoding region. We estimated D' (the LD expressed as a proportion of the maximum disequilibrium possible for a given set of allele frequencies) and r^2 (the pairwise correlation between sites) separately for the three subpopulations corresponding to major geographic subdivisions of the human species (Africans, Europeans, and Asians). At both *IL1B* and *IL10* loci, D' among all three sites was 1 (the maximum possible value). At the *IL1B* locus, r^2 between rs16944 (located at -1,060) and rs1143627 (located at -580) was high in all subpopulations: 0.70 in Africans, 1.00 in Europeans, and 1.00 in Asians. For the other two pairs of SNP sites in the 5' noncoding region of *IL1B*, the r^2 values were low (0.19–0.23) in Africans but higher in Europeans (0.84) and Asians (0.78). By contrast, the pairwise r^2 values among the three pairs of SNPs in the 5' noncoding region of *IL10* were higher in Africans (0.50 in all comparisons) than in Europeans (0.28) or Asians (0.18).

Discussion

Analysis of gene diversity (heterozygosity) at 4,911 SNP sites at 481 loci showed a distinctive patterns depending on the functional effects of the polymorphism. As in previous analyses (Hughes et al. 2003, 2005), nonsynonymous SNPs causing radical amino acid changes showed reduced gene diversity in comparison to silent SNPs in the same genes. This reduction in gene diversity provides evidence that purifying selection is acting to eliminate radical nonsynonymous SNPs, which are presumably disruptive to protein function. There was also evidence of reduced gene

diversity at nonsynonymous SNP sites causing conservative amino acid changes and at SNP sites in the 5' noncoding region, suggesting that these also may often be subject to purifying selection.

By contrast, SNPs at 65 loci with immune system function did not show the overall reduction in gene diversity at conservative nonsynonymous SNP sites or at SNP sites in the 5' noncoding region that was seen at other loci. These results suggest that there has been a relaxation of purifying selection within the human population at immune system loci, parallel to the relaxation of purifying selection at immune system loci seen in comparisons between mammalian species (Murphy 1993, Hughes 1997). This relaxation of purifying selection may in some cases be due simply to reduced functional constraint on immune system proteins in comparison to other proteins (Hughes 1997). On the other hand, there may be at least some cases in the present data set where polymorphisms at immune system loci are maintained by balancing selection. Such balanced polymorphisms may include both nonsynonymous polymorphisms in coding regions and polymorphisms in regulatory elements in 5' noncoding regions.

The hypothesis of balancing selection seems most plausible in the case of a locus with several polymorphic sites, all showing high gene diversity in each of the SNP500 subpopulations. Such sites may include those subject to balancing selection and linked sites hitchhiking along with selected sites. One example involves three SNPs in the 5' noncoding region of *IL1B* at positions -2,022 (rs1143623), -1,060 (rs16944), and -580 (rs1143627). At these three sites, both alleles were found at intermediate frequencies across all subpopulations. Furthermore, the three sites were in LD, and there was a high r^2 between the two most downstream sites. LD within the *IL1A*–*IL1B*–*IL1RN* complex has previously been reported (Smith et al. 2004). Similarly, there were three high-diversity SNP sites in the 5' noncoding region of *IL10*: -1,116 (rs1800896), -835 (rs1800871), and -626 (rs1800872), also in LD. Selectively maintained polymorphisms in the 5' noncoding region presumably related to differences in gene expression and in the case of immune system signaling molecules, such as *IL1B* and *IL10* infectious disease agents, are likely to be responsible for such selection.

Our results show that examining patterns of polymorphism at nucleotide sites of different functional categories can yield important insights into past natural selection. Importantly, because our analyses involve comparisons among SNPs of different functional categories assayed for the same set of genes, our conclusions are not dependent on either the availability of all existing SNPs in the genes under study or on the use of a random sample of the human population. We show that simple statistical analyses of gene diversity at SNP loci can yield insights that suggest candidate loci for further detailed studies regarding the population processes underlying human genetic polymorphism.

Acknowledgements This project has been funded in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract no. NO1-CO-12400. Partial support was provided by N.I.H. grant GM43940 to A.L.H.

References

- Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263–265
- Freudenberg-Hua Y, Freudenberg J, Kluck N, Cichon S, Propping P, Nöthen MM (2003) Single nucleotide variation analysis in 65 candidate genes for CNS disorders in a representative sample of the European population. *Genome Res* 13:2271–2276
- Hughes AL (1997) Rapid evolution of immunoglobulin superfamily domains expressed in immune system cells. *Mol Biol Evol* 14:1–5
- Hughes AL (1999a) Evolutionary diversification of the mammalian defensins. *Cell Mol Life Sci* 56:94–103
- Hughes AL (1999b) Adaptive evolution of genes and genomes. Oxford University Press, New York
- Hughes AL (2002) Evolution of the human killer cell inhibitory receptor family. *Mol Phylogenet Evol* 25:330–340
- Hughes AL, Nei M (1988) Pattern of nucleotide substitution at MHC class I loci reveals overdominant selection. *Nature* 335: 167–170
- Hughes AL, Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proc Natl Acad Sci U S A* 86:958–962
- Hughes AL, Yeager M (1998) Natural selection at major histocompatibility complex loci of vertebrates. *Annu Rev Genet* 32:415–435
- Hughes AL, Packer B, Welch R, Bergen AW, Chanock SJ, Yeager M (2003) Widespread purifying selection at polymorphic sites in human protein-coding loci. *Proc Natl Acad Sci U S A* 100:15754–15757
- Hughes AL, Packer B, Welch R, Bergen AW, Chanock SJ, Yeager M (2005) Effects of natural selection on inter-population divergence at polymorphic sites in human protein-coding loci. *Genetics* (in press)
- Miyata T, Miyazawa S, Yasunaga T (1979) Two types of amino acid substitution in protein evolution. *J Mol Evol* 12:219–236
- Murphy PM (1993) Molecular mimicry and the generation of host defense protein diversity. *Cell* 72:823–826
- Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York
- Packer B, Yeager M, Staats B, Welch R, Crenshaw A, Kiley M, Eckert A, Beerman M, Miller E, Bergen A, Rothman N, Strausberg R, Chanock SJ (2004) SNP500Cancer: a public resource for sequence validation and assay development for genetic variation in candidate genes. *Nucleic Acids Res* 32: D528–D532
- Sherry ST, Warc M-H, Kholodov M, Baler J, Smigelski EM, Sirotkin K (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 29:308–311
- Smith AJP, Keen LJ, Bilingham MJ, Perry MJ, Elson CJ, Kirwan JR, Sims JE, Doherty M, Spector TD, Bidwell JL (2004) Extended haplotypes and linkage disequilibrium in the *IL1R1–IL1A–IL1B–IL1RN* gene cluster: association with knee osteoarthritis. *Genes Immun* 5:451–460
- Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68:978–989
- Sunyaev S, Kondrashov FA, Bork P, Ramensky V (2003) Impact of selection, mutation rate and genetic drift on human genetic variation. *Hum Mol Genet* 15:3325–3330
- Tanaka T, Nei M (1989) Positive Darwinian selection observed at the variable-region genes of immunoglobulin. *Mol Biol Evol* 6:447–459
- Zhao Z, Fu Y-X, Hewett-Emmett D, Boerwinkle E (2003) Investigating single nucleotide polymorphism (SNP) density in the human genome and its implications for molecular evolution. *Gene* 312:207–213